Supplemental Appendix

1. Two-Group Comparisons

Suppose that we record the occurrences of K types of events (e.g., SARS-CoV-2 infection, symptomatic COVID-19, severe COVID-19) for a total of n subjects. For $k = 1, \ldots, K$ and $i = 1, \ldots, n$, let Y_{ki} denote the number of the kth type of event experienced by the ith subject, and T_{ki} denote the corresponding follow-up time. For $i = 1, \ldots, n$, let X_i indicate, by the values 1 versus 0, whether the ith subject receives vaccine or placebo. We assume that Y_{ki} follows a Poisson distribution with mean $\mu_k T_{ki} r_k^{X_i}$, where μ_k is the event rate (per unit time interval) for placebo, and r_k is the rate ratio (i.e., relative risk) between vaccine and placebo. The vaccine efficacy on the kth type of event is defined by $\mathrm{VE}_k = 1 - r_k$, which is the proportionate reduction in cases among the vaccinated persons.

The likelihood for (μ_k, r_k) takes the form

$$\prod_{i=1}^{n} \left(\mu_k T_{ki} r_k^{X_i}\right)^{Y_{ki}} \exp\left(-\mu_k T_{ki} r_k^{X_i}\right).$$

The maximum likelihood estimator of r_k is

$$\frac{\sum_{i=1}^n Y_{ki} X_i / \sum_{i=1}^n T_{ki} X_i}{\sum_{i=1}^n Y_{ki}(1 - X_i) / \sum_{i=1}^n T_{ki}(1 - X_i)}.$$

The score statistic for testing the null hypothesis $H_k : r_k \geq r_{k0}$, i.e., $\mathrm{VE}_k \leq 1 - r_{k0}$, against the alternative hypothesis that $r_k < r_{k0}$, i.e., $\mathrm{VE}_k > 1 - r_{k0}$, takes the form

$$U_k = \sum_{i=1}^{n} (Y_{ki} - \widehat{\mu}_k T_{ki} r_{k0}^{X_i}) X_i,$$

where $\widehat{\mu}_k = \sum_{i=1}^n Y_{ki} / \sum_{i=1}^n T_{ki} r_{k0}^{X_i}$. For large n, the vector of score statistics $(U_1, \ldots, U_K)$ is K-variate zero-mean normal with covariance matrix $\{V_{kl}; k, l = 1, \ldots, K\}$, where

$$V_{kl} = \sum_{i=1}^{n} (Y_{ki} - \widehat{\mu}_k T_{ki} r_{k0}^{X_i})(X_i - \widehat{m}_k)(Y_{li} - \widehat{\mu}_l T_{li} r_{l0}^{X_i})(X_i - \widehat{m}_l),$$

and $\widehat{m}_k = \sum_{i=1}^n T_{ki} r_{k0}^{X_i} X_i / \sum_{i=1}^n T_{ki} r_{k0}^{X_i}$.

For $k = 1, \ldots, K$, we test the null hypothesis $H_k : r_k \geq r_{k0}$ by using the Z-score: $Z_k = U_k / V_{kk}^{1/2}$, which is standard normal under the null hypothesis. We propose to test the

K null hypotheses, adjusting the critical value so as to control the overall type I error at α. Specifically, we reject H_k if the observed value of Z_k is less than the constant c that satisfies the equation

$$\Pr(Z_1 > c, \ldots, Z_K > c) = 1 - \alpha,$$

where $(Z_1, \ldots, Z_K)$ is zero-mean K-variate normal with covariance matrix $\{\rho_{kl}; k, l = 1, \ldots, K\}$, and $\rho_{kl} = V_{kl}/(V_{kk}V_{ll})^{1/2}$. We refer to this method as multiple testing.

To determine which types of events the vaccine is effective against, we adopt a sequential testing procedure, which is more powerful than the above multiple testing method. Let z_k^* be the kth smallest observed value of the Z_k's, and let $(Z_1^*, \ldots, Z_K^*)$ be a zero-mean K-variate normal random vector with a covariance matrix obtained by rearranging the rows and columns of $\{\rho_{kl}; k, l = 1, \ldots, K\}$ according to the order $(z_1^*, \ldots, z_K^*)$. In addition, let $(H_1^*, \ldots, H_K^*)$ be the ordered sequence of the H_k's according to the order $(z_1^*, \ldots, z_K^*)$. Starting with H_1^*, we reject H_k^* $(k = 1, \ldots, K)$ if

$$\Pr(\min_{k \le j \le K} Z_j^* \le z_k^*) \le \alpha,$$

provided that $H_1^*, \ldots, H_{k-1}^*$ have been tested and rejected. It can be shown that the Type I error probability of this procedure is α for any combination of the true H_k's.[1]

We also propose to test the global null hypothesis of no worthwhile vaccine benefit on any endpoint, i.e., $H_0 : r_k \ge r_{k0}$ for all $k = 1, \ldots, K$, by combining the evidence of the vaccine effects on the K endpoints. Specifically, we form a new test statistic by summing the K score statistics and dividing the sum by its standard error:

$$S = \frac{\sum_{k=1}^K U_k}{\left(\sum_{k=1}^K \sum_{l=1}^K V_{kl}\right)^{1/2}}.$$

We refer to S as the combined score test, which is standard normal under H_0.

2. Interim Analyses

Suppose that we perform interim analyses at times $t_1 < t_2 < \ldots < t_M$. Let $U_k(t)$ be the score statistic U_k based on the data collected up to time t. Write $U(t) = \sum_{k=1}^K U_k(t)$. By

the multivariate central limit theorem, the random vector $\{U(t_1), \ldots, U(t_M)\}$ is M-variate normal. Under the Poisson assumption, for $s > t$, the covariance between $U_k(s)$ and $U_l(t)$ is equal to the covariance between $U_k(t)$ and $U_l(t)$. If follows that the covariance between $U(s)$ and $U(t)$ is equal to the variance of $U(t)$. Because of this independent increment property, all standard methods for interim analyses, such as group sequential tests and stochastic curtailment,[2–5] are applicable to the combined test, as well as the individual tests.

3. Simulation Studies

We assigned 27,000 subjects to vaccine or placebo at a ratio of 1:1. For each subject in the placebo group, we generated the time from randomization to infection, the time from infection to disease, and the time from disease to severe disease from the exponential distributions (Fig. 1) with means $\xi\lambda_1$, $\xi\lambda_2$, and $\xi\lambda_3$, respectively, where ξ is a subject-specific random effect that has a gamma distribution with mean 1 and variance 0.5. In the first set of simulation studies, we generated the follow-up time from the Uniform (120,180) distribution. We chose λ_1, λ_2, and λ_3 to yield event proportions of 1% for infection, 0.6% for disease, and 0.12% for severe disease over the 6-month follow-up period. For each subject in the vaccine group, we generated the three event times and the follow-up time in the above manner but chose λ_1, λ_2, and λ_3 to yield the desired values of VE_I, VE_D, and VE_S. To assess statistical power, we set VE_D to 0.6, VE_I to 0.4, 0.5, 0.55 or 0.6, and VE_S to 0.6, 0.7, 0.8 or 0.9. In the second set of simulation studies, we generated the follow-up time from the Uniform (300,360) distribution and chose the event proportions of 2% for infection, 1.2% for disease, and 0.24% for severe disease over the 12-month follow-up period. We set VE_D to 0.3, VE_I to 0.1, 0.2, 0.25 or 0.3, and VE_S to 0.3, 0.4, 0.5 or 0.6.

We evaluated a total of 12 methods: (1) Z_1 alone; (2) Z_2 alone; (3) Z_3 alone; (4) combining U_1 and U_2; (5) combining U_2 and U_3; (6) combining U_1, U_2, and U_3; (7) multiple testing with Z_1 and Z_2; (8) multiple testing with Z_2 and Z_3; (9) multiple testing with Z_1, Z_2, and Z_3; (10) Bonferroni correction for Z_1 and Z_2; (11) Bonferroni correction for Z_2 and Z_3; and (12) Bonferroni correction for Z_1, Z_2, and Z_3. For each method, we performed a one-sided test

with the nominal significance level of 2.5% for the null hypothesis that the vaccine efficacy is at most 0.3 in the first set of simulation studies and is at most zero in the second set of simulation studies. We estimated the power for the 12 methods by simulating 100,000 datasets for each of the 16 combinations of VE_I, VE_D, and VE_S. We considered Z_2 alone the benchmark since most vaccine trials have adopted symptomatic COVID-19 as the primary endpoint.

We implemented the Poisson regression approach described in the previous section, as well as a Cox regression analog.[1,6] The results of the two approaches are almost identical. Here, we report only the Poisson regression results. Poisson regression has clear advantages over Cox regression because it is computationally simple and does not require knowing the event time, but rather whether or not the subject has developed the event of interest by the end of follow-up.

4. General Regression

We consider general Poisson regression, which can be used to perform point and interval estimation, to compare multiple vaccines, and to accommodate baseline risk factors (e.g., age, gender, race, occupation, co-morbidities). As in Appendix 1, there are K types of events and a total n study subjects. For $k = 1, \ldots, K$ and $i = 1, \ldots, n$, let Y_{ki} denote the number of the kth type of event experienced by the ith subject, T_{ki} denote the corresponding follow-up time, and X_{ki} denote a set of covariates (i.e., vaccine indicators, baseline risk factors). (We allow risk factors to depend on the event type.) We assume that Y_{ki} follows a Poisson distribution with mean $\mu_k T_{ki} e^{\beta_k' X_{ki}}$, where μ_k is the baseline event rate (per unit time interval), and β_k is a set of log relative risks.

The likelihood for (μ_k, β_k) takes the form

$$\prod_{i=1}^{n} \left(\mu_k T_{ki} e^{\beta_k' X_{ki}}\right)^{Y_{ki}} \exp\left(-\mu_k T_{ki} e^{\beta_k' X_{ki}}\right).$$

The profile-likelihood score function for β_k is

$$U_k(\beta_k) = \sum_{i=1}^{n} Y_{ki} \left\{X_{ki} - \overline{X}_k(\beta_k)\right\},$$

where $\overline{X}_k(\beta_k) = \sum_{i=1}^n T_{ki} e^{\beta_k' X_{ki}} X_{ki} / \sum_{i=1}^n T_{ki} e^{\beta_k' X_{ki}}$. We obtain the maximum likelihood estimator for β_k, denoted by $\widehat{\beta}_k$, by solving the score equation $U_k(\beta_k) = 0$ via the Newton-Raphson algorithm. We then estimate μ_k by $\widehat{\mu}_k = \sum_{i=1}^n Y_{ki} / \sum_{i=1}^n T_{ki} e^{\widehat{\beta}_k' X_{ki}}$.

For large n, the estimators $\widehat{\beta}_k$ $(k = 1, \ldots, K)$ are jointly normal with means β_k $(k = 1, \ldots, K)$. In addition, the covariance matrix between $\widehat{\beta}_k$ and $\widehat{\beta}_l$ is $A_k^{-1} V_{kl} A_l$, where $A_k = \partial U(\widehat{\beta}_k)/\partial \beta_k$, and

$$V_{kl} = \sum_{i=1}^{n} (Y_{ki} - \widehat{\mu}_k T_{ki} e^{\widehat{\beta}_k' X_{ki}})(Y_{li} - \widehat{\mu}_l T_{li} e^{\widehat{\beta}_l' X_{li}}) \big\{ X_{ki} - \overline{X}_k(\widehat{\beta}_k) \big\} \big\{ X_{li} - \overline{X}_l(\widehat{\beta}_l) \big\}'.$$

Let η_k be the component of β_k that corresponds to the vaccine effect on the kth endpoint. We extract the maximum likelihood estimator $\widehat{\eta}_k$ from $\widehat{\beta}_k$ and extract the covariance matrix of $(\widehat{\eta}_1, \ldots, \widehat{\eta}_K)$, denoted by $\Sigma = \{\sigma_{kl}; k, l = 1, \ldots, K\}$, from the covariance matrices between $\widehat{\beta}_k$ and $\widehat{\beta}_l$ $(k, l = 1, \ldots, K)$. For $k = 1, \ldots, K$, we calculate the Z-score for testing the null hypothesis $H_k : \eta_k \geq \eta_{k0}$ by using the Z-score: $Z_k = (\widehat{\eta}_k - \eta_{k0})/\sigma_{kk}^{1/2}$. We can then use these Z-scores for the multiple testing and sequential testing procedures described in Appendix 1.

Suppose that $\eta_1 = \eta_2 = \ldots = \eta_K = \eta$. Then we can estimate η by the weighted linear combination: $\widehat{\eta} = \sum_{k=1}^K w_k \widehat{\eta}_k$, where $(w_1, \ldots, w_K)' = (e' \Sigma^{-1} e)^{-1} \Sigma^{-1} e$, and $e = (1, \ldots, 1)'$.[1] In addition, we test the global null hypothesis $H_0 : \eta_k \geq \eta_{k0}$ for all $k = 1, \ldots, K$ by the standard-normal statistic:

$$S = \frac{\sum_{k=1}^K w_k \widehat{\eta}_k}{\big(\sum_{k=1}^K \sum_{l=1}^K w_k w_l \sigma_{kl} \big)^{1/2}},$$

which is analogous to the combined score test given at the end of Appendix 1. Although the assumption of a common vaccine effect on the K endpoints may not hold, $\widehat{\eta}$ provides a concise summarization of the vaccine effects, and S provides a valid test of overall vaccine efficacy.

References

1. Wei LJ, Lin DY, Weissfeld L (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. J Am Stat Assoc, 84: 1065-1073.

2. O'Brien PC, Fleming TR (1979). A multiple testing procedure for clinical trials. Biometrics, 35: 549-556.

3. Lan KKG, DeMets DL (1983). Discrete sequential boundaries for clinical trials. Biometrika, 70: 659-663.

4. Lan KKG, Wittes J (1988). The *B*-value: a tool for monitoring data. Biometrics, 44: 579-585.

5. Jennison C, Turnbull BW (1999). Group Sequential Methods With Applications to Clinical Trials. CRC Press.

6. Lin DY, Zeng D, Eron JJ (2020). Evaluating the efficacy of therapies in COVID-19 patients. Clinical Infectious Diseases, https://doi.org/10.1093/cid/ciaa1231.

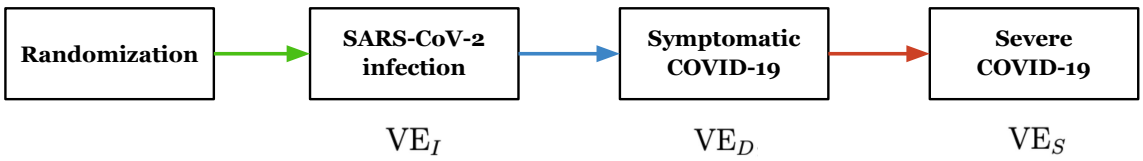

Figure 1. A 4-state model for a Phase 3 COVID-19 vaccine trial. The time between two adjacent events follows an exponential distribution, with different rates between vaccine and placebo to achieve vaccine efficacy of VE_I, VE_D, and VE_S for infection, disease, and severe disease, respectively.